Disponível em www.bad.pt/publicacoes

PAPER



A b-on e a preservação digital de periódicos eletrónicos: o projeto Portico

*Teresa Costa^a**^aFCT/FCCN, Portugal, teresa.costa@fccn.pt*

Resumo

O número de periódicos científicos em formato digital tem vindo a aumentar significativamente na última década. Existem cada vez mais conteúdos a transitar para o digital ou até mesmo a nascer em formato digital, não tendo, por isso, uma cópia “física”. No mundo do papel, a responsabilidade da preservação estava vinculada ao proprietário e era tradicionalmente uma função das bibliotecas e dos arquivos. Na era digital, o vínculo entre o proprietário, a preservação e o acesso foi alterado.

A escala e a complexidade da infraestrutura tecnológica exigida, os conhecimentos especializados e os processos de controlo de qualidade necessários à preservação de recursos eletrónicos são extremamente exigentes e dispendiosos.

Tendo a biblioteca do conhecimento online (b-on) recursos exclusivamente eletrónicos esta é uma questão fundamental, pelo que foram analisadas várias soluções tendo a escolha recaído sobre a solução Portico da JSTOR.

Palavras-chave: Biblioteca do Conhecimento Online, b-on, periódicos eletrónicos, preservação digital, arquivo, Portico

Introdução

Os periódicos científicos têm desempenhado um papel fundamental no sistema de comunicação científica. Com a migração do impresso para o ambiente digital, a manutenção do acesso à informação científica do passado, passou a ser uma questão fundamental, pelo que a preservação dos periódicos científicos eletrónicos se tornou numa das principais preocupações dos bibliotecários.

Muitos dos *stakeholders* da publicação científica perceberam a importância e urgência de tomarem medidas que garantissem a preservação e o acesso a este tipo de conteúdo de forma eficaz.

Sendo a b-on *e-only* tem naturalmente como preocupação a manutenção do acesso aos conteúdos subscritos por parte de todas as suas instituições, mesmo quando as mesmas por questões diversas se vêm obrigadas a deixar o consórcio.

«Com o aumento da produção de informação em formato digital, tem sido questionada cada vez mais a importância de se ter garantida a sua disponibilização e preservação por longos períodos de tempo [...]. Os objetos digitais não podem ser deixados em formatos obsoletos para serem transferidos depois de longos períodos de

negligência para repositórios digitais» (Márdeder Arellano, 2004, p.16).

Ora, a escala e a complexidade da infraestrutura tecnológica exigida, os conhecimentos especializados e os processos de controlo de qualidade necessários à preservação de recursos eletrónicos são extremamente exigentes e dispendiosas. A preservação dos conteúdos digitais apresenta uma série de questões inerentes à sua condição de digital que têm de ser consideradas, como por exemplo, a obsolescência dos formatos, a gestão de copyright, a manutenção do acesso, a integridade dos dados, o controle de qualidade, etc. Isso acontece porque a informação digital depende, na sua mais pura essência, de meios técnicos para ser acedida e, sobretudo, corretamente interpretada. Este aparato tecnológico de intermediação – formado por *hardware*, *software*, formatos - está em constante mutação, tendo ciclos de obsolescência cada vez mais rápidos determinados principalmente pelo binómio inovação e competição.

Tendo a b-on recursos exclusivamente eletrónicos esta é, pois, uma questão fundamental, uma vez que existe a preocupação de garantir a preservação dos conteúdos subscritos por forma a:

- Proteger o investimento assegurando o acesso aos conteúdos em caso de perda por motivos de força maior;
- Assegurar o acesso de conteúdos subscritos aquando da saída de membros; e
- Assegurar o acesso a títulos transferidos.

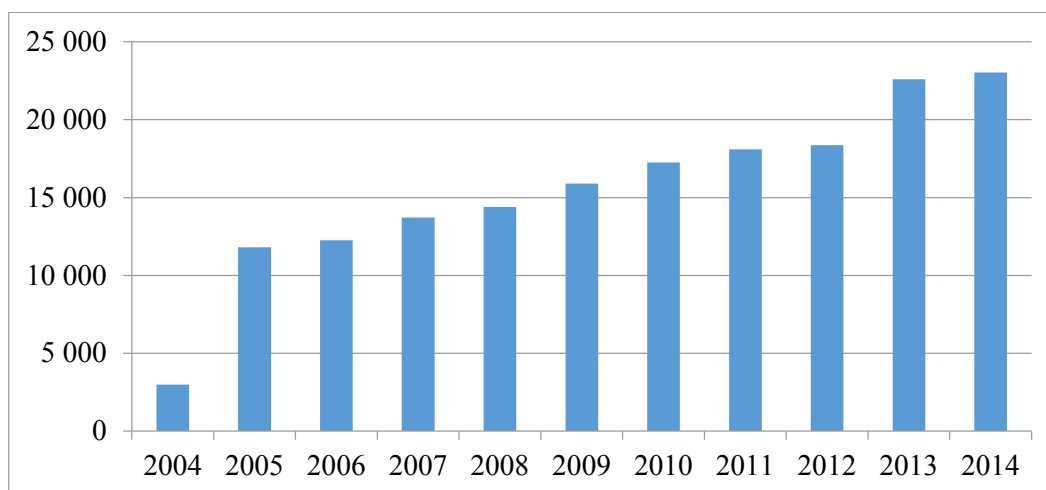


Gráfico 1 – Evolução do número de títulos subscritos

O número de títulos subscritos pela b-on tem vindo a aumentar ao longo dos anos, como pode ser observado no Gráfico 1, pelo que a manutenção do acesso aos mesmos é uma prioridade. Muitos dos editores garantem o acesso perpétuo às publicações subscritas como parte dos contratos e respeitando os anos de duração dos mesmos. Este é um modelo no qual as editoras se comprometem a manter o acesso online às coleções/títulos subscritos às instituições que por alguma razão se viram forçadas a deixar de ser membros b-on. Este modelo pode ser visto como um compromisso por parte das editoras para manter o acesso aos conteúdos contratados. Tem a vantagem de que os custos de gestão, manutenção e atualização estão centrados no editor. Além disso, tem o benefício adicional da familiaridade que o utilizador tem com plataformas dos editores e a facilidade para colocar *links* para os recursos no OPAC através do uso da etiqueta 856 do MARC (*Machine Readable Cataloging Record*). Assim, o

utilizador pode aceder através do catálogo local ou diretamente a partir da interface do editor. As desvantagens desse modelo estão relacionados com a perda do acesso, uma vez que a sua manutenção depende exclusivamente dos editores e o mesmo pode deixar de ocorrer pelos mais diversos motivos (desaparecimento da editora, fusão da mesma com outra editoras,...). Porém, nem todas as editoras disponibilizam sem custos a manutenção do acesso perpétuo e, assim, permitir o acesso aos conteúdos subscritos. Daí que seja necessário arranjar e assegurar outras alternativas de forma a garantir o acesso aos conteúdos eletrónicos a longo prazo.

Método

Assim, e de modo a garantir o acesso perpétuo aos conteúdos eletrónicos tornou-se necessária a existência de serviços de arquivo compatíveis com o formato digital. Estes serviços de preservação digital são garantidos e mantidos por entidades independentes das editoras. São serviços fornecidos por terceiros, mediante pagamento de uma taxa anual que permitem o acesso aos títulos subscritos em caso de falha. São serviços em que os editores concordam em fornecer autorização de preservação dos seus conteúdos em servidores externos com a possibilidade de dar acesso à informação em caso de falha dos serviços. Destas soluções destacam-se três modelos que são atualmente utilizadas em todo o mundo quer por consórcios ou instituições individualmente e que são o Portico, o LOCKSS e o CLOCKSS.

A b-on fez uma análise das três soluções tendo para tal estabelecido como essenciais os seguintes critérios:

- Termos e condições de acesso
- Suporte à tipologia consórcio e/ou instituição
- Conteúdos (editores, títulos e cobertura temporal)
- Custos (adesão e suporte e infraestrutura técnica)

Resultados

Vejamos uma breve síntese das características de cada uma das soluções.

LOCKSS (Lots of Copies Keep Stuff Safe)

Visão geral

Fundada pelas Stanford University Libraries, em 1998, com o apoio da National Science Foundation, da Mellon Foundation, da Soros Foundation e da Biblioteca do Congresso.

Permite às bibliotecas participantes recolher, armazenar, preservar e fornecer o acesso a conteúdos que tenham subscrito ou adquirido.

Como funciona

Fornece às bibliotecas *software open source* de modo a permitir a instalação de uma LOCKSS Box com as seguintes funções:

- Recolher o conteúdo dos sites de destino através um *web crawler* semelhante ao utilizado pelos motores de pesquisa

- Comparar o conteúdo obtido, com o mesmo conteúdo recolhido por outras LOCKSS Boxes, corrigindo diferenças.
 - Atuar como web proxy ou cache, garantindo o acesso aos conteúdos dos editores ou aos conteúdos preservados. Também pode dar acesso aos conteúdos através de metadados (*Open URLs*)
 - Fornecer uma interface que permite aos bibliotecários identificar novos títulos a preservar, monitorar o estado dos títulos preservados e controlar o acesso aos mesmos.

Quando é ativado o acesso

O acesso é ativado sempre que (e por qualquer motivo) o conteúdo não pode ser visualizado no *website* do editor (ou intermediário). O acesso pós cancelamento da subscrição do contrato com o editor (*Post-cancellation access* – PCA) é suportado.

Vantagens

Esta solução tem as seguintes vantagens:

- Permite que as bibliotecas possam exercer controlo sobre os conteúdos licenciados (como acontece com as subscrições em papel)
- As bibliotecas têm garantia de *post-cancellation access* (com autorização do editor), e acesso ao arquivo, sempre que haja problemas de comunicação com os servidores dos editores o número de editoras e de títulos disponíveis está a aumentar
- Procura duplicar/replicar os conteúdos de tal forma que mesmo havendo uma catástrofe se mantenha assegurado o acesso aos conteúdos protegidos por parte das instituições participantes.
- Requer investimentos relativamente pequenos em termos de pessoal e equipamentos, mas as bibliotecas precisam monitorar o servidor local e garantir que o mesmo esteja a funcionar corretamente.
- Conta com mais de 15 anos de experiência

Desvantagens

- Algumas das grandes editoras STM não participam no LOCKSS devido a preocupações sobre a segurança dos conteúdos licenciados.
- Algumas editoras, como a Elsevier, não permitem a inclusão dos seus conteúdos com receio da perda de controlo dos mesmos que ficam distribuídos por várias “caixas” LOCKSS.
- Por vezes, nem todos os títulos de uma editora ou todos os números de um título em particular estão incluídos

CLOCKSS (Controlled LOCKSS)

Visão geral

Lançado em 2006, é baseado numa colaboração sem fins lucrativos entre bibliotecas e editores. É um “*dark archive*” baseado no *software* LOCKSS em que um número limitado de bibliotecas assume o papel de arquivo em nome de uma comunidade mais ampla.

O arquivo é gerido pelas partes interessadas (editores e bibliotecários têm a mesma palavra na

decisão sobre procedimentos, prioridades e ativação de conteúdos).

Como funciona

O sistema CLOCKSS preserva os ficheiros da fonte e o documento, fornecidos pelos editores. Várias instituições a nível mundial assumem a responsabilidade de arquivo tendo cada uma dois servidores para criação de uma rede de “*dark repositories*”.

Quando é acionado o acesso

Quando os conteúdos não estão disponíveis a partir dos editores passam a estar disponíveis sob uma licença *Creative Commons*. O *Post-cancellation* não é suportado.

Vantagens

Principais vantagens:

- Relação próxima com a comunidade, na qual um pequeno número de bibliotecas assume a responsabilidade de arquivo trabalhando em parceria com os editores.
- Inclui alguns dos maiores editores STM que não participam no LOCKSS devido a preocupações com a segurança dos conteúdos licenciados.
- É uma opção de custo relativamente baixo, especialmente no arquivo em *outsourcing*

Desvantagens

- O *Post-cancellation access* não é suportado
- Os custos a longo prazo e de sustentabilidade do projeto não são claros
- Acesso a conteúdos não disponíveis através das editoras participantes é feito gratuitamente para todos - não apenas para as bibliotecas membro

PORTICO

Visão geral

Lançado em 2005 com o suporte da JSTOR (Ithaka), da Biblioteca do Congresso e da Mellon Foundation, o Portico é um serviço, sem fins lucrativos, de preservação de conteúdos digitais.

Como funciona

Os ficheiros fornecidos pelos editores são “normalizados” num formato padrão de arquivo. Enquanto o “look and feel” do original é perdido, os arquivos são geridos de forma fiável sendo actualizados à medida que a tecnologia e os formatos de arquivo se forem alterando (evitando ficar obsoletos).

Quando é acionado o acesso

O acesso é acionado se o conteúdo deixa de estar disponível aquando das seguintes situações:

- Se a instituição cancela a subscrição
- Se editor desaparece,
- Se o título deixa de ser publicado,
- Se as edições anteriores não estão mais disponíveis ou se se verifica uma catástrofe.

Vantagens

- Suporta o *post-cancellation* (o Portico assegura o acesso após cancelamento de mais de 87% dos periódicos eletrónicos e 100% dos ebooks).
- Retira às bibliotecas o encargo de administrar os conteúdos da biblioteca e fornece a garantia de preservação a longo prazo. No entanto, como CLOCKSS, exige uma mudança cultural para as bibliotecas, em termos de seu papel tradicional de “guardião”.
- O número de editoras participantes tem aumentado bastante e algumas das principais editoras STM aderiram à iniciativa.
- Os ficheiros para arquivo fornecidos pelos editores são "normalizados" para um formato padrão de arquivo que vai sendo atualizado.
- Fornece um modelo de negócio justo, quer para as editoras quer para as bibliotecas. No entanto, alguns veem na dependência dos editores para obtenção de receita uma fraqueza e as taxas anuais para muitas bibliotecas podem ser elevadas.

Analisaram-se as três soluções e após avaliadas as vantagens e desvantagens de cada uma delas a escolha recaiu sobre o Portico (<http://www.portico.org>). O Portico é o resultado de uma iniciativa do JSTOR (acrónimo para Journal Storage) que em 2002 procurou encontrar e desenvolver um modelo de preservação para a literatura científica (periódicos e livros). Atualmente é um serviço autónomo que recebe apoios da Biblioteca do Congresso dos EUA, do JSTOR, do Ithaka e da Andrew W. Mellon Foundation.

A estratégia consiste na centralização do processo de preservação no Portico, retirando às bibliotecas e aos editores a tarefa de se ocuparem autonomamente desse problema. Em contrapartida, é-lhes pedido apoio financeiro e, no caso dos editores, autorização para que o Portico possa preservar e dar acesso permanente aos recursos, mesmo nos casos em que o editor desaparece, descontinua um título ou em que a biblioteca deixa de subscrever o serviço ao editor.

T tecnicamente, o processo resume-se à receção do material original e à respetiva conversão para o formato de preservação. Os formatos, esquemas de metadados e modelos usados são: Journal Archiving and Interchange DTD, PREMIS, OAIS.

A missão do Portico é preservar a literatura científica/académica publicada em formato eletrónico e assegurar que estas publicações permaneçam acessíveis no futuro. Tanto editores como bibliotecas cooperam com o Portico na preservação de publicações; a abordagem da comunidade base do Portico assenta na colaboração ativa entre estes intervenientes-chave.

O Portico assegura o acesso aos conteúdos arquivados quando ocorrerem situações específicas (*trigger event access*), ou quando os títulos já não estiverem disponíveis no editor ou noutra fonte, designadamente:

- O editor desaparece ou cessa atividade,
- O título deixa de ser publicado,
- As edições anteriores deixam de estar disponíveis ou
- Se se verifica uma falha catastrófica e continuada da plataforma do editor
- A instituição cancela a subscrição e o editor delegou no Portico o *Post Cancellation*

Access.

Este último ponto é de extrema importância e como refere Fenton (2006, p.83)

“In addition to these trigger events, both publishers and libraries have recognized that, in some cases, even after a library has terminated a license to an electronic resource, it may be necessary for that library to continue to have ongoing access. This is commonly known as “perpetual access” or post-cancellation access. A publisher may choose to extend perpetual access to a library and that access can be provided through the Portico archive if the publisher desires. In addition, select librarians at participating libraries are granted password-controlled access to the archive for verification purposes. This verification access, which is granted to the entire archive, is not intended to be used as a replacement for commercial document delivery services or to fulfill interlibrary loan requests. Finally, all publishers participating in the archive have full access to their own content and any content for which a trigger event prevails”.

Discussão

O arquivo de periódicos científicos eletrônicos é uma questão complexa e com consequências sociais, económicas, legais, organizacionais e técnicas “ Preservation of digital materials is a dynamic and evolving process: the methods are changing, as are the technical requirements” (Ross, 2012,p 47). Apesar de existirem atualmente diversas soluções, não existe nenhuma que garanta a preservação de todos os conteúdos devido à natureza e tipologia dos mesmos, pelo que ainda não há soluções ideais.

O CLOCKSS não se adequa às necessidades de preservação da b-on, uma vez que não suporta o *post-cancellation*; o LOCKSS garante o acesso em tempo real quando o conteúdo não está disponível, no entanto, não integra os principais editores da b-on, pelo que o Portico surgiu como a solução mais viável, uma vez que apresenta a maior taxa de cobertura dos conteúdos subscritos e garante o *post-cancellation* para a maioria dos títulos.

No caso concreto da b-on, muitos dos conteúdos subscritos incluem uma elevada taxa de títulos disponibilizados por um agregador, a EBSCO, e que por isso não são passíveis de preservar, no entanto, e das várias soluções analisadas o Portico parece ser a que melhor satisfaz os objetivos de preservação e a que melhor se adequa aos requisitos do projeto de preservação. Assim, e desde 2012 que a b-on tem um contrato com o Portico.

Conclusões

A preservação digital é hoje um grande desafio, ao qual a b-on não está indiferente. É, pois, necessário prever e garantir a disponibilidade a longo prazo dos conteúdos científicos digitais, pelo que é fundamental encontrar uma solução de arquivo adequada. São várias as abordagens tendo cada modelo as suas próprias forças e fraquezas. Não existe, no momento, uma solução que garanta a preservação de tudo o que é subscrito. O importante é conhecer as várias soluções existentes, avaliá-las e acompanhá-las de modo a tomar a decisão mais conveniente e que melhor responda às necessidades e anseios da comunidade.

É importante reconhecer que a preservação de revistas eletrónicas é um tipo de seguro, e não é por si só uma forma de acesso. A preservação é uma maneira de gerir o risco, em primeiro lugar, contra a perda permanente do acesso aos conteúdos *online* e, em segundo, contra ter o acesso diário interrompido por um período prolongado após uma falha por parte editor.

Cabe, então, à b-on continuar a acompanhar o desenvolvimento das iniciativas de

preservação, estabelecer parcerias internacionais visando soluções de armazenamento local por forma a partilhar serviços e custos e continuar a pressionar os editores a participar nos projetos de preservação e facilitar o acesso aos conteúdos.

Como a tecnologia da informação está em constante mudança e a um ritmo acelerado, muitas mais iniciativas podem ser desenvolvidas, havendo assim muito trabalho ainda a ser feito de modo a alcançar um serviço de arquivo ideal para os periódicos eletrónicos.

Referências bibliográficas

CLOCKSS (Controlled LOCKSS). [Em linha]. [Consult. 21 Jun. 2015]. Disponível na Internet:< URL: <https://www.clockss.org/>>

FENTON, Eileen Gifford. “An overview of Portico: an electronic archiving service”. *Serials Review*. [Em linha]. 32:2 (2006) 81 – 86.[Consult. 22.02.2015]. Disponível na Internet <URL: <http://www.sciencedirect.com/science/article/pii/S0098791306000384>>

LOTS OF COPIES KEEP STUFF SAFE, LOCKSS. [Em linha]. [Consult. 21 Jun. 2015]. Disponível na Internet:< URL: <http://lockss.org/>>

MÁRDERO ARELLANO, Miguel – “Preservação de documentos digitais”. *Ciência da Informação*. [Em linha]. 33:2 (2004) 15-27. [Consult. 22.02.2015]. Disponível na Internet <URL: <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/305/1452>>

PORTICO. [Em linha]. [Consult. 21 Jun. 2015]. Disponível na Internet:< URL: <http://www.portico.org/>>

ROSS, Seamus - “Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries”. *New Review of Information*. [Em linha]. 17:1 (2012) 43-68. [Consult. a 22.02.2015]. Disponível na Internet <URL: <http://dx.doi.org/10.1080/13614576.2012.679446>>